REMARKS

Claim 17 and 19-28 are currently pending. Claim 17 is the sole independent claim. Claims 18 and 29 have been cancelled.

In the Office Action it was stated that claim 29 is directed to an independent or distinct invention. Claim 29 has been cancelled without traverse.

On page 4 of the Office Action, the information disclosure statement filed June 1, 2004 was cited as failing to comply with the provisions of 37 CFR 1.97, 1.98 and MPEP §609 because the NCBI accession numbers failed to include an author and date. This has been addressed in a Supplemental IDS filed herewith.

Claims 17 and 21-28 were rejected under 35 USC 112, first paragraph, for failing to comply with the written description requirement. The Office Action states that no support could be found for the amended claims that recite "wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID NO:6 have **at least 92% identity** based on the Clustal alignment method". In order to advance prosecution, Applicants have amended claim 17 to incorporate the limitation of dependent claim 18, now cancelled, which is directed to sequences with "at least **95%** identitity." As a result, this rejection no longer applies and Applicants respectfully request withdrawal of this rejection under 35 USC 112, first paragraph.

Claims 17-18 and 21-28 were rejected under 35 USC 112, first paragraph, as for failing to comply with the written description requirement.

The specification discloses to one of ordinary skill in the art a representative number of cycloartenol synthases with at least 95% sequence identity to SEQ ID NO:6. The specification at page 6, lines 8-19, discloses alterations in nucleotide sequence (SEQ ID NO:6, for example) that are not expected to alter functionality, such as alterations that produce a chemically equivalent amino acid at a given site or alterations in the N- or C-terminal portions. Thus, from the foregoing, the skilled artisan would understand the specification to disclose a representative number of polynucleotide sequences that have different nucleotide substitutions but are within 95% sequence identity of SEQ ID NO:6 and encode functional cycloartenol synthases.

Submitted herewith is a copy of Abe and Prestwich, PNAS 92: 9274-9278 (1995) (Abe). Abe discloses the molecular cloning, characterization, and functional expression of a rat oxidosqualene cyclase (OSC). Abe states on page 9275, second

column, rat OSC shares the highest homology with the cycloartenol synthase from *Arabidopsis thaliana*. The high homology is consistent with the similarity in the reactions catalyzed by the two enzymes (OSC and cycloartenol synthase). Abe states that the two enzymes share a very ancient ancestry. See page 9274, column 2. Abe goes on to report conserved structural features of OSC, including the QW motif [(K/R)(G/A)XX(F/Y/W)(L/I/V)XXXQXXXGXW], which is highly conserved repetitive motif, rich in aromatic amino acids (page 9276, column 1). The authors believe that the QW motif plays a structural or functional role in catalysis by stabilizing the carbocationic intermediates. Abe also disclose a highly conserved DCTAEA motif on page 9277, column 1. This motif is well conserved in all known OSCs and has been hypothesized to play an important role in the cyclase activity based on a point mutation in the corresponding motif in the bacterial enzyme almost completely abolishing activity.

Abe also reports on page 9277, second column, the successful complementation between the *Arabidopsis* cycloartenol synthase and the yeast *erg7* (lanosterol synthase gene) mutant strain. This result confirms the expectation that cycloartenol synthases from divergent species are functionally similar.

Attached hereto as Appendix A is an alignment of SEQ ID NO:6 with the *Glycyrrhiza glabra* and *Arabidopsis thaliana* cycloartenol synthases and the OSC from *Rattus norvegicus* and *Saccaromyces cerevisiae* (NCBI General Identifier Nos. 4589852, 541855, 1098634 and 1169548, respectively). Amino acids conserved among all sequences are indicated with an asterix above the conserved residues. Dashes are used by the program to maximize alignment of the sequences. The QW motifs and the highly conserved DCTAEA motif as disclosed by Abe are shown in the alignment. The comparison among the sequences in Appendix A demonstrates that the soybean cycloartenol synthase (SEQ ID NO:6) possesses the above-described motifs that are characteristic of cycloartenol synthase activity.

In view of the above discussion, Applicants believe sufficient relevant identifying characteristics have been disclosed to allow one skilled in the art to determine Applicants were in possession of the claimed invention at the time of filing current application.In view of the foregoing, Applicants respectfully request reconsideration and withdrawal of this Section 112, first paragraph rejection.

Claims 17-28 were rejected under 35 USC 112, first paragraph, for failing to comply with the enablement requirement.

Applicants believe that above discussion concerning written description, incorporated herein by reference, equally obviates this rejection. Accordingly, Applicants respectfully request reconsideration and withdrawal of this rejection under 35 USC 112, first paragraph.

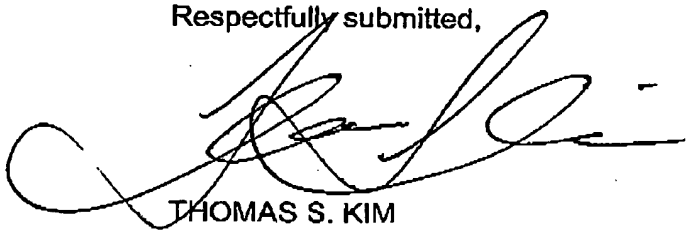
Claim 21 was rejected under 35 USC 101 because the claimed invention was directed to non-statutory matter.

Claim 21 has been amended and the phrase "A cell comprising" has been replaced with "An isolated cell" as suggested in the Office Action. Thus, withdrawal of the rejection is respectfully requested.

It is respectfully submitted that all pending claims are in condition for allowance which allowance is respectfully requested. Copies of the Abe reference, Appendix A, Supplemental Information Disclosure, Petition for Two-Month Extension, and appropriate fees accompany this response.

Please charge any fees or credit any overpayment of fees which are required in connection with the filing of this Response to Deposit Account No.: 04-1928 (E. I. du Pont de Nemours and Company).

Respectfully submitted,



THOMAS S. KIM
AGENT FOR APPLICANTS
REGISTRATION NO.: 51,009
TELEPHONE: 302-992-4061
FACSIMILE: 302-892-1026

Date: 5/4/06

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 9274–9278, September 1995
Biochemistry

Molecular cloning, characterization, and functional expression of rat oxidosqualene cyclase cDNA

(cholesterol biosynthesis/cDNA cloning/active site/QW motif)

IKURO ABE AND GLENN D. PRESTWICH*

Department of Chemistry, State University of New York, Stony Brook, NY 11794-3400

Communicated by William J. Lennarz, State University of New York, Stony Brook, NY, July 3, 1995 (received for review May 10, 1995)

ABSTRACT A cDNA encoding rat oxidosqualene lanosterol-cyclase [lanosterol synthase; (*S*)-2,3-epoxysqualene mutase (cyclizing, lanosterol-forming), EC 5.4.99.7] was cloned and sequenced by a combination of PCR amplification, using primers based on internal amino acid sequence of the purified enzyme, and cDNA library screening by oligonucleotide hybridization. An open reading frame of 2199 bp encodes a M_r 83,321 protein with 733 amino acids. The deduced amino acid sequence of the rat enzyme showed significant homology to the known oxidosqualene cyclases (OSCs) from yeast and plant (39–44% identity) and still retained 17–26% identity to two bacterial squalene cyclases (EC 5.4.99.-). Like other cyclases, the rat enzyme is rich in aromatic amino acids and contains five so-called QW motifs, highly conserved regions with a repetitive β-strand turn motif. The binding site sequence for the 29-methylidene-2,3-oxidosqualene (29-MOS), a mechanism-based irreversible inhibitor specific for the vertebrate cyclase, is well-conserved in all known OSCs. The hydropathy plot revealed a rather hydrophilic N-terminal region and the absence of a hydrophobic signal peptide. Unexpectedly, this microsomal membrane-associated enzyme showed no clearly delineated transmembrane domain. A full-length cDNA was constructed and subcloned into a pYEUra3 plasmid, selected in *Escherichia coli* cells, and used to transform the OSC-deficient uracil-auxotrophic SGL9 strain of *Saccharomyces cerevisiae*. The recombinant rat OSC expressed was efficiently labeled by the mechanism-based inhibitor [^{3}H]29-MOS.

Oxidosqualene lanosterol-cyclase (OSC) [lanosterol synthase; (*S*)-2,3-epoxysqualene mutase (cyclizing, lanosterol-forming), EC 5.4.99.7] catalyzes the conversion of (3*S*)-2,3-oxidosqualene to lanosterol, forming a total of six new carbon–carbon bonds in a single reaction (1). The regulation of OSC levels *in vivo* has clinical importance and has been a potential target for the design of hypercholesteremic drugs (2). The formation of lanosterol is initiated in the chair-boat-chair-like conformation of oxidosqualene, and the proton-initiated cyclization is postulated to proceed through a series of rigidly held carbocationic intermediates (1). The intermediate C-20 protosterol cation then undergoes backbone rearrangement to yield the lanosterol skeleton (Fig. 1) (3–6). Oxidosqualene is also the precursor for a variety of other polycyclic triterpenes, and the relationship between enzyme structure and cyclization mechanism has been extensively studied (1).

Several OSCs have been purified to homogeneity from vertebrate (7–9), plant (10–13), and yeast sources (14). Recently, two OSC enzymes have been cloned and sequenced from the fungi *Saccharomyces cerevisiae* (15, 16) and *Candida albicans* (17–19). An oxidosqualene cycloartenol-cyclase was cloned and sequenced from the plant *Arabidopsis thaliana* (20). In addition, two bacterial squalene hopene-cyclases (SCs) (EC 5.4.99.-) have been cloned from *Alicyclobacillus acidocaldarius* (21) and *Zymomonas mobilis* (22). The predicted molecular masses of OSCs ranged from 80 to 90 kDa and the deduced amino acid sequences showed that the bacterial, fungal, and plant enzymes share a very ancient ancestry. A highly conserved repetitive β-strand turn motif rich in aromatic amino acids (the QW motif) occurs in all OSCs and SCs, and this likely serves a structural or catalytic role in the cyclization reaction (23, 24). We previously reported mapping of the active site of rat liver OSC using 29-[^{3}H]methylidene-2,3-oxidosqualene ([^{3}H]29-MOS), a mechanism-based irreversible inactivator specific for vertebrate OSCs (8, 25–27). We describe here the cDNA cloning and characterization of rat liver OSC, and we report the expression of a functional rat OSC in the sterol auxotrophic SGL9 yeast strain.†

EXPERIMENTAL PROCEDURES

Amino Acid Analysis and Protein Sequencing of Rat OSC. Purified rat liver OSC was labeled with [^{3}H]29-MOS and was digested with CNBr or with endoproteinase Lys-C and sequenced by Edman degradation as described (26). The following partial sequences were obtained: CNBr 8-kDa fragment, VRYLRSVQLPDGGWGLHHEDKSTVFG (aa 129–154); Lys-C 50-kDa fragment, NNVCPDDMY (aa 279–287); CNBr 6-kDa fragment, HKGGFPFSTLDDGWIVADDTAEALKAVLLLQE (aa 439–470).

Library Construction and Screening. Male Sprague–Dawley rats were fed with cholestyramine and Fluvastatin (hydroxymethylglutaryl-CoA reductase inhibitor) for 1 week to induce 15-fold higher OSC activity relative to control rats (28). From the induced liver, total RNA was obtained by the acid guanidium thiocyanate/phenol/chloroform extraction method (29). A cDNA library was then constructed from 5 μg of poly(A)$^+$ mRNA with the ZAP cDNA synthesis kit with Lambda Uni-ZAP XR vector and the Gigapack II packaging extract (Stratagene). The titer of the resulting library was 7.5 × 10^5 plaque-forming units/μl after one amplification.

For screening, a PCR amplified 500-bp cDNA fragment obtained as described below was labeled with [α-^{32}P]dATP. Recombinant plaques (≈1 × 10^6) were transferred to nylon membranes (Hybond-N; Amersham) and hybridized with the radiolabeled probe in 5× SSPE (0.75 M NaCl/0.05 M NaH_2PO_4/5 mM EDTA, pH 7.4)/5× Denhardt's reagent/50% formamide/0.1% SDS/100 μg of salmon testes DNA per ml at 42°C for 24 hr. Membranes were washed three times with 0.1× SSC (15 mM NaCl/1.5 mM sodium citrate, pH 7.0)/0.1% SDS at 55°C for 20 min. Two positive phage clones were purified and subcloned into Bluescript II SK− phagemids. One of the positive clones contained an ≈2-kbp cDNA insertion.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SC, squalene cyclase; OSC, oxidosqualene cyclase; 29-MOS, 29-methylidene-2,3-oxidosqualene.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. U31352).

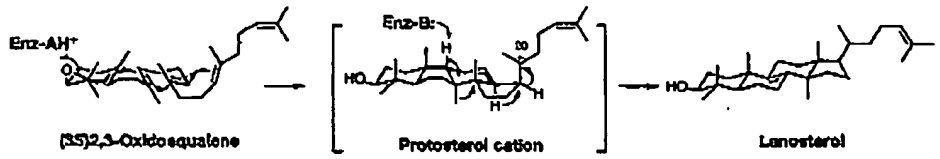


FIG. 1. Formation of lanosterol from (3S)-2,3-oxidosqualene by OSC.

Both strands of cDNA were completely sequenced by the dideoxynucleotide chain-termination method using Sequenase version 2.0 (United States Biochemical) (30).

PCR Screening of cDNA Library. Degenerate oligonucleotides 5′-TG(T/C)CCNGA(T/C)GA(T/C)ATGTA-3′ (sense) coding CPDDMY of the Lys-C 50-kDa fragment (aa 282–287) and 5′-TCNGCNAC(G/A/T)ATCCANCC-3′ (antisense) coding GWIVAD of the CNBr 6-kDa fragment (aa 451–456) were used for amplification of a partial sequence of the cDNA (524 bp). After phage particles were disrupted (70°C for 5 min and then 4°C), 30 cycles of PCR were carried out as follows: 1 min at 94°C, 2 min at 45°C, and 3 min at 72°C. The gel-purified PCR products were ligated into pGEM-T plasmid (Promega) and sequenced.

To obtain a full-length sequence, PCR amplification of the 5′ end of the cDNA from the library was first performed using three primer sets of gene-specific internal antisense primers and nonspecific T3 primer coding part of the λ ZAP vector arm (primer ratio, 10:1). Primer sets amplified 1000-, 570-, and 200-bp fragments, each with the same 5′ end sequence and each missing several 5′ nucleotides required for an expected full-length cDNA.

To obtain the full 5′ sequence, the 5′ inverse PCR method was used (31, 32). Poly(A)$^+$ mRNA (10 μg) was first reverse-transcribed by avian myeloblastosis virus reverse transcriptase (200 units), using 50 pmol of gene-specific antisense primer 5′-ATGGCCACTGCACCACCTT-3′ (nt 551–569) instead of the oligo(dT) primer. The reaction was run at 50°C for 1 hr to minimize the secondary structure of the RNA. After second-strand DNA synthesis, the cDNA was circularized by T4 DNA ligase (30 units) at 15°C for 50 hr. PCR was then performed by using primers 5′-AAGTCCATCTCTGCCGAC-3′ (antisense, nt 95–112) and 5′-CGTGTCACATAGCACAC-3′ (sense, nt 335–351) to amplify a 388-bp DNA fragment containing the 5′ end sequence and an additional 41 bp of 5′ noncoding sequence.

Northern Blot Analysis. Poly(A)$^+$ mRNA from rat liver was size-fractionated by electrophoresis on a 0.8% agarose gel containing 2.2 M formaldehyde. After the gel was treated with 0.05 M NaOH, the denatured RNA was transferred to nylon membrane and hybridized with a ^{32}P-labeled cDNA probe (using the 5′ end 200-bp PCR product) and autoradiographed as described above.

Expression of Recombinant Rat OSC in Yeast. A yeast/*E. coli* shuttle vector pYEUra3 (Clontech) was used for functional expression of the rat OSC cDNA in the yeast OSC-deficient mutant strain SGL9 (*erg7 ura3-52 hem3-6 gal2*) that was obtained from J. H. Griffin (16). The strain SGL9 is a segregant of a cross between OSC-deficient mutant GL7 strain (*erg7-s hem3-6 gal2*) (33) and strain 9a (*ura3-52*). A full-length cDNA (2833 bp) was obtained by ligating the N-terminal PCR fragment (nt 1–1310) and the C-terminal fragment (nt 1311–2833) excised from the above described 2-kbp Bluescript insertion by *Nde* I/*Xho* I digestion. Here the PCR primers used are 5′-AAGGATCC ATG ACC GAG GGC ACG TGT CTG C-3′ (sense) (the *Bam*HI site is underlined) and 5′-G GAA ACC ACC CTT GTG CAT ATG GCG-3′ (antisense) (the *Nde* I site is underlined). The amplified DNA was digested with *Bam*HI/*Nde* I prior to the ligation reaction. The cDNA was cloned into the *Bam*HI/*Xho* I site of the plasmid pYEUra3 downstream of the yeast *GAL1* promoter. After transformation of *Escherichia coli* DH5α cell and selection on LB/ampicillin plates, plasmids were recovered and checked for the insertion. The SGL9 strain was then transformed with the recovered plasmid by the LiAc method (34), and transformants were selected on synthetic complete (SDC) medium without uracil and ergosterol.

Enzyme Assay and Affinity Labeling. The Ura$^+$ Erg$^+$ yeast transformant cell was first grown in 50 ml of SDC medium containing 2% glucose. When cells grew to OD_{600} = 0.5, the medium was replaced with the SDC medium containing 2% galactose and incubated for another 3 hr to induce expression of the OSC gene under the *GAL1* promoter. The cell pellet (100 mg) obtained after centrifugation was washed once with 100 mM Tris·HCl (pH 7.4), resuspended in 200 μl of 100 mM Tris·HCl (pH 7.4) containing 1 mM dithiothreitol, and then broken by Vortex mixing with glass beads (eight times for 45 s). The cell-free extracts (100 μl) were incubated with (3S)2,3-[^{14}C]oxidosqualene (50 μM; 5.5 mCi/mmol; 1 Ci = 37 GBq) or [^{3}H](3S)29-MOS (1 μM; 1.8 Ci/mmol) as described at 30°C for 17 hr (8). For control experiments, the SGL9 strain alone and the SGL9 strain transformed with pYEUra3 plasmid lacking the rat OSC insert were used.

RESULTS AND DISCUSSION

Rat OSC cDNA was cloned and sequenced from a rat liver cDNA library by a combination of PCR amplification based on partial amino acid sequence of the purified enzyme followed by cDNA library screening with labeled oligonucleotides for hybridization. The 2874-bp cDNA contained a 41-bp 5′ noncoding region, a 2199-bp open reading frame encoding a M_r 83,321 protein of 733 amino acids, and 634 bp of 3′ noncoding region (Fig. 2). A polyadenylylation signal (AATAAA) precedes the poly(A) sequence by 27 bp, and an in-frame stop codon is located 36 bp upstream of the start codon. The predicted molecular mass of 83 kDa is consistent with the molecular size of the purified protein (78 kDa) estimated by SDS/PAGE (8). In addition, the sequence (underlined) surrounding the start codon (boldface), GCTGTCATGA showed good homology to the Kozak consensus sequence, GCC(G/A)CCATG(G/A), for the 5′ noncoding sequence of vertebrate mRNA (35). Finally, on Northern blot analysis of rat liver poly(A)$^+$ mRNA, a single transcript of ≈3 kb was detected (data not shown).

The deduced amino acid sequence showed significant similarity to the yeast and plant OSCs: 40.2% identity (295/733) with *S. cerevisiae* OSC, 39.0% identity (286/733) with *C. albicans* OSC, and 44.2% identity (324/733) with *A. thaliana* OSC (cycloartenol synthase). It is interesting that rat lanosterol synthase showed highest similarity to plant cycloartenol synthase. The postulated cyclization mechanism of oxidosqualene to cycloartenol is essentially the same as that for lanosterol formation, except for the final 9β,19-cyclopropane ring closure instead of C-9 proton elimination (1). Only a slight modification of the active site of the enzyme could determine the product specificity. Rat OSC also showed substantial identity to two prokaryotic SCs that directly cyclize squalene into the pentacyclic triterpene hopene: 26.3% identity (193/733) with *Z. mobilis* SC and 16.6% identity (122/733) with thermoaci-

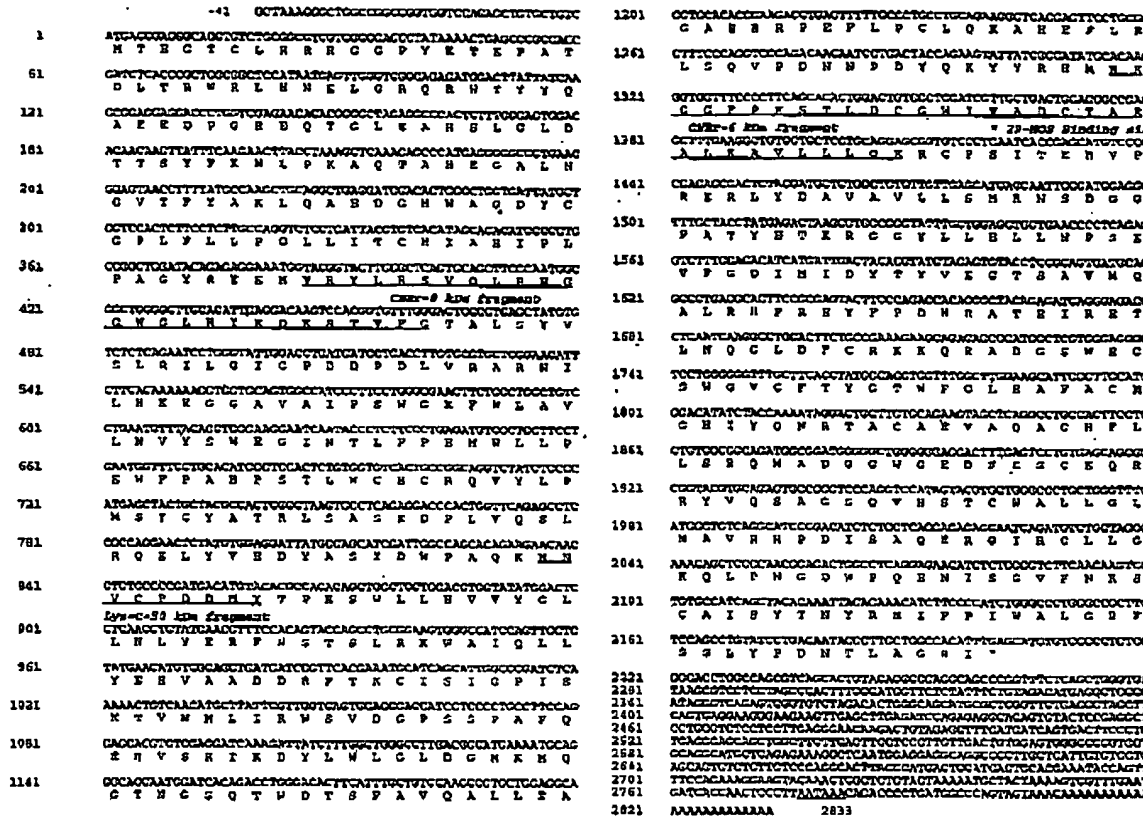


FIG. 2. Nucleotide and predicted amino acid sequence of rat liver OSC. A 2199-bp open reading frame encoding 733 aa is shown with the single-letter code used for the translated amino acids. Nucleotide numbering begins with the ATG start codon. Amino acid sequences of peptides obtained from CNBr and Lys-C cleavages are underlined.

dophilic *A. acidocaldarius* SC. In general, the sequence is more highly conserved in the C-terminal region than in the N terminus. Except for these cyclases, no other significant sequence similarity was obtained from the GenBank/EMBL data bases.

From sequence comparisons of eukaryotic OSCs and bacterial SCs, we have previously reported the existence of the QW motif (23), a highly conserved repetitive motif rich in aromatic amino acids: [(K/R)(G/A)XX(F/Y/W)(L/I/V)XXXQXXXGXW]. There are six repeats of the QW motif in rat OSC, four in the C-terminal one-third and two in the N-terminal one-third of the protein (Fig. 3). The motif was well-conserved in both eukaryotic OSCs and prokaryotic SCs. According to the PEPPLOT program (36), a typical QW motif contains part of a β-strand at the N terminus and a turn at the C terminus. We (37) and others (16) have postulated that the aromatic amino acids of the QW motif might play a structural or functional role in catalysis by stabilizing the carbocationic intermediates through cation–π interactions.

Of 733 amino acids of rat OSC, 175 residues (24%) are completely conserved in all four known OSCs (rat, yeast, *Candida*, plant). Overall, rat OSC contains a disproportionately higher number of aromatic amino acid residues that are completely conserved (16): Phe (13 of 28 residues), Trp (13 of 24 residues), and Tyr (13 of 35 residues). The negatively charged Asp and Glu residues in rat OSC are less highly conserved; Asp (6 of 36, 1 at the 29-MOS binding site) and Glu (10 of 43). As proposed by Johnson and colleagues (38–41), negative point charges at the active site of the enzyme could control the course of the cyclization reaction by stabilization of the developing cationic centers on the cyclizing substrate. The electron density required for stabilization could arise from anionic or from aromatic residues. The higher conservation of aromatic residues relative to anionic residues is noteworthy in view of the suggestion that cation–π interactions may stabilize cyclization intermediates (16, 23). Interestingly, an anionic residue, D-456, is implicated in stabilization of the C-20 cation after tetracyclization but prior to hydride methyl migrations (26, 27). Furthermore, there are six conserved Cys (C-282, C-457, C-534, C-585, C-617, C-701, one at the 29-MOS binding site), and four conserved His (H-145, H-226, H-233, H-290) residues. The presence of an essential cysteinyl group in the active site of the enzyme has been previously suggested, since the OSC activity can be efficiently inhibited by SH reagents such as *p*-chloromercuribenzenesulfonic acid and *N*-ethylmaleimide (13, 42, 43). In contrast, diethyl pyrocarbonate, a histidyl-selective reagent, does not inhibit OSC activity (42, 43). Finally, a disproportionately higher number of Gly (29 of 59 residues) and Pro (11 of 38 residues) are also conserved, suggesting important conserved elements of secondary structure.

According to Kyte–Doolittle hydropathy plot analysis (44), rat OSC is a moderately hydrophilic protein (Fig. 4). It has a hydrophilic region at the N terminus (D-5–E-53) as found in the yeast and plant cyclases. Furthermore, no signal peptide sequence was observed at the N terminus. Surprisingly, there were no significantly hydrophobic regions that may serve as possible membrane-spanning regions. Indeed, the EMBL neural network program (45, 46) predicted the absence of helical transmembrane domains at the 95% confidence level. The

QW-1	Rat	673	RGIRCLLGKQLPNGDW
	Yeast	675	RGIDLLKNBQESSGEW
	Candida	671	RGIQFLNKRQLPTGEW
	Plant	702	RAARYLINAQMENGDF
	Bacteria-a	572	RGVQYLVETQRPDGGW
	Bacteria-z	595	KGIRWLAQNQDEEGLW
QW-2	Rat	615	QACHFLLSRQMADGGW
	Yeast	617	KGCDFLVSKQMKDGGW
	Candida	613	KGCDFLISKQLPDGGW
	Plant	640	KACEFLLSKQQPSGGW
	Bacteria-a	614	KALDWVEQHQNPDGGW
	Bacteria-z	536	KAVAWLKTIQNEDGGW
QW-3	Rat	563	QGLDFCRKKQRADGSW
	Yeast	568	IAIEFIKKSQLPDGSW
	Candida	563	SAIQYILDSQDNIDGSW
	Plant	591	KAVKFIESIQAADGSW
	Bacteria-a	466	RAVEYLKREQKPDGSW
	Bacteria-z	488	AAVDYLLKEQEEDGSW
QW-4	Rat	485	DAVAVLLSMRNSDGGF
	Yeast	488	EGIDVLLNLQN-IGSF
	Candida	482	DAVEVLLQIQN-VGEW
	Plant	514	EAVNVIISLQNADGGL
	Bacteria-a	398	KGFRWIVGHQSSNGGW
	Bacteria-z	420	RAMEWTICMQSDNGCW
QW-5	Bacteria-a	332	KAGEWLLDRQITVPGDW
	Bacteria-z	350	SALSWLKPQQILDVKCDW
QW-6	Rat	127	EMVRTLRSVQLPNGGW
	Yeast	127	ELIRTIVNTAHPVDGGW
	Candida	129	EMIRTIVNTAHPVDGGW
	Plant	149	EMRRYLYNHQMEDGGW
QW-7	Rat	80	NGVTPYAKLQAEDGHW
	Yeast	79	NGASPFKLLQEPDSGIF
	Candida	72	KGADPLKLLQLQNGIF
	Plant	99	RGLDPYSTIQAHDGHW
	Bacteria-a	17	RAVEYLLSCQKDEGYW
	Bacteria-z	24	KATRALLEKQQQDGHW

FIG. 3. Summary of the highly conserved repetitive QW motifs [(K/R)(G/A)XX(F/Y/W)(L/I/V)XXXQXXXGXW] in OSC from six species: rat, *Rattus rattus* OSC; yeast, *S. cerevisiae* OSC; Candida, *C. albicans* OSC; plant, *A. thaliana* OSC (cycloartenol synthase); bacteria-a, *A. acidocaldarius* SC; bacteria-z, *Z. mobilis* SC. Frequently occurring residues are in boldface; hyphens indicate gaps introduced to maximize alignment.

microsomal cyclases appear to bind loosely to membranes; this may explain the ease of solubilization under mild conditions. A similar situation was reported for the recently cloned rat squalene epoxidase (47), another membrane-associated enzyme that catalyzes formation of 2,3-oxidosqualene, the substrate for OSC. Furthermore, in rat OSC, there are five possible N-glycosylation sites (N-383, N-517, N-606, N-692, and N-698). Although preliminary results showed evidence for protein glycosylation, incubation with N-glycosidase did not reduce OSC activity (unpublished results).

We reported previously that the two adjacent Asp residues (D-456 and D-457) in a putative DDTAEA motif (see Fig. 2) were labeled with the mechanism-based irreversible inhibitor [^{3}H]29-MOS (26, 27). However, according to the deduced amino acid sequence, it now appears that the 29-MOS binding site sequence just N-terminal of the QW-4 motif is actually DCTAEA, instead of DDTAEA. Interestingly, this sequence is well-conserved in all the known OSCs (15, 16) despite the fact that neither yeast nor plant OSCs can be labeled with [^{3}H]29-MOS (8). In Edman sequencing of the labeled peptide, elution of the first steroid-modified Asp phenylthiohydantoin derivative might have carried over to the next cycle, leading to miscalling of the residue. This carryover would be particularly difficult to detect given that it was followed by an underivatized Cys. Therefore, only the Asp residue of the DCTAEA motif was actually labeled with the suicide substrate. Site-directed mutagenesis experiments will provide a further test for this hypothesis. Recently, Poralla and co-workers reported that a point mutation of the first Asp residue of the corresponding

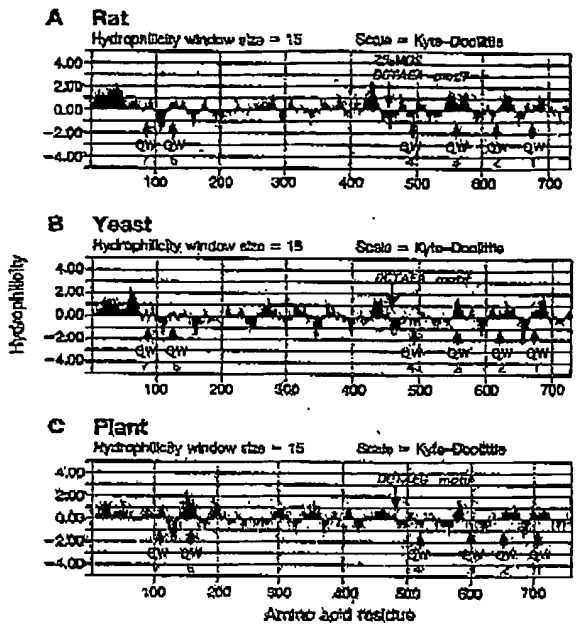


FIG. 4. Hydropathy plots for rat OSC (*A*), yeast *S. cerevisiae* OSC (*B*), and plant *A. thaliana* cycloartenol synthase OSC. (*C*) Deduced amino acid sequence was analyzed by MACVECTOR 4.1 sequence analysis software (Kodak). The QW motifs and the 29-MOS binding site (DCTAEA) are indicated. Plant and yeast OSCs are not labeled by [^{3}H]29-MOS (8).

DDTAVV motif of bacterial *A. acidocaldarius* SC caused almost complete loss of cyclase activity.‡

The 2.8-kbp rat OSC cDNA was cloned into the yeast/*E. coli* shuttle vector pYEUra3 (Clontech) and functionally expressed in the OSC-deficient yeast mutant strain SGL9 (*erg7 ura3-52 hem3-6 gal2*) (16) under the control of the *GAL1* promoter. Recent cloning of yeast and plant OSCs (15–20) were all achieved by complementation of this OSC-deficient (*erg7*) mutant strain (33). The low copy number centromeric plasmid pYEUra3 is a modified version of λYES (pSE937) (48), which was successfully used for expression of yeast OSC in the SGL9 strain by Griffin and co-workers (16). The *GAL1* promoter is induced >1000-fold by the presence of galactose but is repressed by the presence of glucose (49). Like other OSC genes, the rat OSC cDNA appeared to complement the *erg7* deficiency of the yeast mutant strain, since the transformant cells could grow on media lacking ergosterol.

Enzyme activity for conversion of (3*S*)-2,3-[^{14}C]oxidosqualene to lanosterol was detected only after galactose induction; the activity (27 fmol of lanosterol per min per mg of cell pellet) was <10% of wild-type yeast cyclase activity. A similar low OSC activity was found for heterologous expression of *Candida* OSC in the yeast *erg7* strain (17). In contrast, plant *A. thaliana* cycloartenol synthase OSC was successfully expressed in the *erg7* strain under the control of *PGK1* promoter (20). The low specific activity of the recombinant rat OSC may be due to decreased stability of the heterologous protein; after galactose induction, no novel protein band was observable at ≈80 kDa on SDS/PAGE.

As a critical test of the authenticity of this heterologously expressed recombinant rat OSC, the cell-free extract was incubated with [^{3}H]29-MOS, a mechanism-based inhibitor that

‡Poralla, K., Ochs, D. & Feil, C., American Oil Chemists' Society Symposium on the Regulation of Biosynthesis and Function of Isopentenoids, May 9–10, 1994, Atlanta.

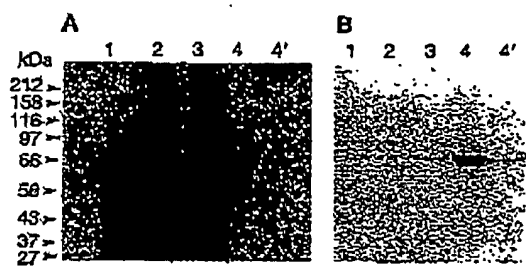


FIG. 5. [^{3}H]29-MOS affinity labeling of cell-free extracts from the transformed yeast mutant strain SGL9 (*erg7*). Rat OSC cDNA was expressed under the control of yeast *GAL1* promoter. (*A*) SDS/PAGE (7.5%) gel stained with Coomassie blue. (*B*) Corresponding fluorogram. Lanes contain cell-free extracts as follows: lane 1, SGL9; lane 2, transformed SGL9 before galactose induction; lane 3, transformed SGL9 after induction; lane 4, purified rat native OSC loaded to visualize stained protein; lane 4', purified rat native OSC loaded with enzyme activity equivalent to that of SGL9 extract in lane 3.

forms a covalent linkage to an Asp residue of the vertebrate but not plant or yeast cyclase active site (8, 26). Fig. 5 shows a single radioactive band with the same molecular size (≈80 kDa) as the native rat OSC. At equivalent levels of OSC enzymatic activity, the intensity of labeling of the yeast-expressed recombinant rat OSC by [^{3}H]29-MOS is essentially identical to that of the native rat enzyme. This report thus constitutes isolation, sequencing, and functional expression of a vertebrate enzyme responsible for cyclization of 2,3-oxidosqualene to lanosterol.

Note. After the cDNA sequencing was completed,[§] a short communication describing the sequence of a cDNA encoding rat liver OSC appeared (50). The communication, which does not provide any experimental details, shows a sequence (GenBank accession no. D45252) that differs from that shown in Fig. 2 by seven amino acid residues; moreover, expression of a functional OSC protein was not reported.

[§]Abe, I. & Prestwich, G. D., Poster Session, 14th Enzyme Mechanism Conference, January 4–7, 1995, Scottsdale, AZ.

We are indebted to Professor John H. Griffin at Stanford University for the gift of yeast mutant strain SGL9 and for his advice on cDNA expression. We thank Mr. Thomas Fischer (Center of Analysis and Synthesis of Macromolecules, Stony Brook) for Edman peptide sequencing and Ms. Brenda A. Madden for synthesis of (3*S*)-2,3-[^{14}C]oxidosqualene and [^{3}H]29-MOS. This work was supported by the Center for Biotechnology/New York State Science and Technology Foundation, Sandoz Research Institute, DuPont/NEN, and the National Institutes of Health (Grant GM 44836).

1. Abe, I., Rohmer, M. & Prestwich, G. D. (1993) *Chem. Rev.* **93**, 2189–2206.
2. Abe, I., Tomesch, J. C., Wattanasin, S. & Prestwich, G. D. (1994) *Nat. Prod. Rep.* **11**, 279–302.
3. Eschenmoser, A., Ruzicka, L., Jeger, O. & Arigoni, D. (1955) *Helv. Chim. Acta* **38**, 1890–1904.
4. van Tamelen, E. E. (1982) *J. Am. Chem. Soc.* **104**, 6480–6481.
5. Corey, E. J., Virgil, S. S. & Sarshar, S. (1991) *J. Am. Chem. Soc.* **113**, 8171–8172.
6. Corey, E. J. & Virgil, S. C. (1991) *J. Am. Chem. Soc.* **113**, 4025–4026.
7. Kusano, M., Abe, I., Sankawa, U. & Ebizuka, Y. (1991) *Chem. Pharm. Bull.* **39**, 239–241.
8. Abe, I., Bai, M., Xiao, X.-y. & Prestwich, G. D. (1992) *Biochem. Biophys. Res. Commun.* **187**, 32–38.
9. Moore, W. R. & Schatzman, G. L. (1992) *J. Biol. Chem.* **267**, 22003–22006.
10. Abe, I., Ebizuka, Y. & Sankawa, U. (1988) *Chem. Pharm. Bull.* **36**, 5031–5034.
11. Abe, I., Sankawa, U. & Ebizuka, Y. (1989) *Chem. Pharm. Bull.* **37**, 536–538.
12. Abe, I., Ebizuka, Y., Seo, S. & Sankawa, U. (1989) *FEBS Lett.* **249**, 100–104.
13. Abe, I., Sankawa, U. & Ebizuka, Y. (1992) *Chem. Pharm. Bull.* **40**, 1755–1760.
14. Corey, E. J. & Matsuda, S. P. T. (1991) *J. Am. Chem. Soc.* **113**, 8172–8174.
15. Corey, E. J., Matsuda, S. P. T. & Bartel, B. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2211–2215.
16. Shi, Z., Buntel, C. & Griffin, J. H. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 7370–7374.
17. Kelly, R., Miller, S. M., Lai, M. H. & Kirsch, D. R. (1990) *Gene* **87**, 177–183.
18. Buntel, C. J. & Griffin, J. H. (1992) *J. Am. Chem. Soc.* **114**, 9711–9713.
19. Roessner, C. A., Min, C., Hardin, S. H., Harris-Haller, L. W., McCollum, J. C. & Scott, A. I. (1993) *Gene* **127**, 149–150.
20. Corey, E. J., Matsuda, S. P. T. & Bartel, B. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11628–11632.
21. Ochs, D., Kaletta, C., Entian, K.-D., Beck-Sickinger, A. & Poralla, K. (1992) *J. Bacteriol.* **174**, 298–302.
22. Reipen, I., Poralla, K., Sahm, H. & Sprenger, G. (1995) *Microbiology (Reading, U.K.)* **141**, 155–161.
23. Poralla, K., Hewelt, A., Prestwich, G. D., Abe, I., Reipen, I. & Sprenger, G. (1994) *Trends Biochem. Sci.* **19**, 157–158.
24. Poralla, K. (1994) *Bioorg. Med. Chem. Lett.* **4**, 285–290.
25. Xiao, X.-y. & Prestwich, G. D. (1991) *J. Am. Chem. Soc.* **13**, 9673–9674.
26. Abe, I. & Prestwich, G. D. (1994) *J. Biol. Chem.* **269**, 802–804.
27. Abe, I. & Prestwich, G. D. (1995) *Lipids* **30**, 231–234.
28. Bai, M. (1991) Ph.D. dissertation (State University of New York, Stony Brook).
29. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
31. Towner, P. & Gärtner, W. (1992) *Nucleic Acids Res.* **20**, 4669–4670.
32. Zeiner, M. & Gehring, U. (1994) *BioTechniques* **17**, 1050–1054.
33. Gollub, E. G., Liu, K.-p., Dayan, J., Adlersberg, M. & Sprinson, D. B. (1977) *J. Biol. Chem.* **252**, 2846–2854.
34. Schiestl, R. H. & Gietz, R. D. (1989) *Curr. Genet.* **16**, 339–346.
35. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
36. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
37. Kumpf, R. A. & Dougherty, D. A. (1993) *Science* **261**, 1708–1711.
38. Johnson, W. S., Lindell, S. D. & Steele, J. (1987) *J. Am. Chem. Soc.* **109**, 5852–5853.
39. Johnson, W. S., Telfer, S. J., Cheng, S. & Schubert, U. (1987) *J. Am. Chem. Soc.* **109**, 2517–2518.
40. Johnson, W. S. (1991) *Tetrahedron* **47**, xi–1.
41. Johnson, W. S., Buchanan, R. A., Bartlett, W. R., Tham, F. S. & Kullnig, R. K. (1993) *J. Am. Chem. Soc.* **115**, 504–515.
42. Duriatti, A. & Schuber, F. (1988) *Biochem. Biophys. Res. Commun.* **151**, 1378–1385.
43. Hoshino, T., Williams, H. J., Chung, Y. & Scott, A. I. (1991) *Tetrahedron* **47**, 5925–5932.
44. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
45. Post, B. & Sander, C. (1994) *Proteins* **19**, 55–72.
46. Post, B., Casadio, R., Fariselli, P. & Sander, C. (1995) *Protein Sci.* **4**, 521–533.
47. Sakakibara, J., Watanabe, R., Kanai, Y. & Ono, T. (1995) *J. Biol. Chem.* **270**, 17–20.
48. Ramer, S. W., Elledge, S. J. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11589–11593.
49. Johnston, M. & Davis, R. W. (1984) *Mol. Cell. Biol.* **4**, 1440–1448.
50. Kusano, M., Shibuya, M., Sankawa, U. & Ebizuka, Y. (1995) *Biol. Pharm. Bull.* **18**, 195–197.

Appendix A

Comparison of the amino acid sequences of the cycloartenol synthase from *Glycine max* (SEQ ID NO:6), *Glycyrrhiza glabra* and *Arabidopsis thaliana* and the oxidosqualene lanosterol-cyclases from *Rattus norvegicus* and *Saccaromyces cerevisiae* set forth in NCBI General Identifier No's. 4589852, 541855, 1098634 and 1169548, respectively. Amino acids conserved among all sequences are indicated with an asterix above the conserved residues. Dashes are used by the program to maximize alignment of the sequences. The conserved motifs as disclosed by Abe and Prestwich [(1995) PNAS 92: 9274-9278] are indicated below as follows: QW motifs are underlined and the DCTAEA motif is indicated by arrows below the conserved residues.

```
             *                                 **  *
SEQ ID NO:6  M------------WKLKFAEGGNPWLRTLN-NHVGRQVWEF-DPKLGSPQDLLEIEKARQ
Gi: 4589852  M------------WKLKIAEGGSPWLRTVN-NHVGRQVWEF-DPKLGSPEDLLEIEKARQ
Gi: 541855   M------------WKLKIAEGGSPWLRTTN-NHVGRQFWEF-DPNLGTPEDLAAVEEARK
Gi: 1098634  MTEGTCLRRRGGPYKTEPATDLTRW--RLH-NELGRQRWTYYQAEEDPGREQTGLEAHSL
Gi: 1169548  MTE-FYSDTIGLP-------KTDPRLWRLRTDELGRESWEYLTPQQAAND----------

                                 *
SEQ ID NO:6  NFHDNRFTHKHSADLLMRMQFARENPTREVLPKVGVKDIEDVTQEIVTKTLRRAVSFHST
Gi: 4589852  NFHDNRFTHKHSADLLMRIHFAKENPMNEVLPKVRVKDIEDVTEETVKTTLRRAINFHST
Gi: 541855   SFPDNRFVQKHSADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLDFYST
Gi: 1098634  GLDTTSY-------------FKN-------LPKAQTAHEGALN-----------GVTPYAK
Gi: 1169548  --PPSTFTQW----LLQDPKFPQPHPER---NKHSPDFSAFD------ACHNGASFFKL
                                                                   QW-7
             **  * *     * ** *   *                     *  **         ***
SEQ ID NO:6  LQCHD-GHWPGDYGGPMFLMPGLVITLSITGALNTVLTEEHRKEICRYLYNHQNK-DGGW
Gi: 4589852  LQSHD-GHWPGDYGGPMFLMPGLVITLSITGALNAVLTEEHRKEICRYLYNHQNK-DGGW
Gi: 541855   IQAHD-GHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNE-DGGW
Gi: 1098634  LQAED-GHWAGDYGGPLFLLPGLLITCHIA---HIPLPAGYREEMVRYLRSVQLP-NGGW
Gi: 1169548  LQEPDSGIFPCQYKGPMFMTIGYVAVNYIAG---IEIPEHERIELIRYIVNTAHPVDGGW
                                                              QW-6
             ***    ** **  * *  *****             *       ***    *** **
SEQ ID NO:6  GLHIEGPSTMFGSVLSYITLRLLGEGPNDGQGEMEKARDWILGHGGATYITSWGKMWLSV
Gi: 4589852  GLHIEGPSTMFGSVLNYVALRLLGEGPNDRQGEMEKGRDWILGHGGATFITSWGKMWLSV
Gi: 541855   GLHIEGPSTMFGSVLNYVTLRLLGEGPNDGDGDMEKGRDWILKYGGATNITSWGKMWLSV
Gi: 1098634  GLHIEDKSTVFGTALSYVSLRILGIGPDDP--DLVRARNILHKKGGAVAIPSWGKPWLAV
Gi: 1169548  GLHSVDKSTVFGTVLNYVILRLLGLPKDHPVCA-KARSTLLRLGGAIGSPHWGKIWLSA

             *    * * *   ** ****   * **   * * * **** **                *
SEQ ID NO:6  LGVYEWSGNNPLPPEIWLLPYMLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPISPTVLSL
Gi: 4589852  LGVYEWSGNNPLPPEIWLLPYVLPIHPGRMWCHCRMVYLPMSYLYGKRFVGPITPTILSL
Gi: 541855   LGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGPITSTVLSL
Gi: 1098634  LNVYSWEGINTLFPEMWLLPEWFPAHPSTLWCHCRQVYLPMSYCYATRLSASEDPLVQSL
Gi: 1169548  LNLYKWEGVNPAPPETWLLPYSLPMHPGRWWVHTRGVYIPVSYLSLVKFSCPMTPLLEEL

             * *                  *     * **          *               *
SEQ ID NO:6  RKELYTVPYHDIDW-DQARNLCAKEDLYYPHPLVQDILWASLHKFLEPILMHWPGKRLRE
Gi: 4589852  RKELYTIPYHDIDW-NQARNLCAKEDLYYPHPLVQDILWASLHKFLEPILMHWPGKRLRE
Gi: 541855   RKELFTVPYHEVNW-NEARNLCAKEDLYYPHPLVQDILWASLHKFLEPILMHWPGKKLRE
Gi: 1098634  RQELYVEDYASIDWPAQKNNVCP-DDMYTPHSWLLHVVYGLLNLYE-----RFHSTSLRK
Gi: 1169548  RNEIYTKPFDKINFSKNRNTVCGVD-LYYPHSTTLNIA-NSLVVFYEKYLRNRFIYSLSK
```

```
                 *   * *        *          * **       ** * **
SEQ ID NO:6  KAIISALEHIHYEDENTRYICIGPVNKVLNMLCCW-VEDPNSEAFKLHLPRIYDYLWIAE
Gi: 4589852  MAIKTAIEHIHYEDDNTRYLCIGPVNKVLNMLCCW-VEDPNSEAFKLHLPRIYDYLWIAE
Gi: 541855   KAIRTAIEHIHYEDENTRYLCIGPVNKVLNMLCCW-VEDPNSEAFKLHLPRIYDYLWIAE
Gi: 1098634  WAIQLLYEHVAADDRFTKCISIGPISKTVNMLIRWSVDGPSSPAFQEHVSRIKDYLWLAE
Gi: 1169548  KKV---YDLIKTELQNTDSLCIAPVNQAFCALVTLIEEGVDSEAFQRLQYRFKDALFHGP

             ** ** ** * ** * *          *       *       *       *
SEQ ID NO:6  DGMKMQGYNGSQLWDTAFAVQAII--ASNLIEEFGPTIRKAHTYIKNSQVLEDC-PGDLN
Gi: 4589852  DGMKMQGYNGSQLWDTAFTAQAII--SSNLIEEYGPTLRKAHTYIKNSQVLEDC-PGDLN
Gi: 541855   DGMKMQGYNGSQLWDTGFAIQAIL--ATNLVEEYGPVLEKAHSFVKNSQVLEDC-PGDLS
Gi: 1098634  DGMKMQGTNGSQTWDTSFAVQALLEAGAHRRPEPLPCLQKAHEFLRLSQVPDNN-P-DYQ
Gi: 1169548  QGMTIMGTNGVQTWDCAFAIQYFFVAGLAERPEFYNTIVSAYKFLCHAQFDTECVPGS-

              **  **  ***  *    ***** ***            *  *    **    *
SEQ ID NO:6  KWYRHISKGAWPFSTGDHGWPISDCTAEGLKAVLLLSKIAP-EIVGEPIDVKRLYDSVNV
Gi: 4589852  KWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAP-EIVGEPLDAKRLYDSVNV
Gi: 541855   YWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKVPK-EIVGEPIDAKRLYEAVNV
Gi: 1098634  KYYRHMHKGGFPFSTLDCGWIVADCTAEALKAVLLLQERCP-SIT-EHVPRERLYDAVAV
Gi: 1169548  --YRDKRKGAWGFSTKTQGYTVADCTAEAIKAIIMVKNSPVFSEVHHMISSERLFEGIDV
                                     ▲▲▲▲▲
                *    *  ****        *  ** * ** *   * *****          *
SEQ ID NO:6  ILSLQNED----GGFATYELKRSYNWLEIINPAETFGDIVIDYPYVECTSAAIQALASFR
Gi: 4589852  ILSLQNED----GGFATYELTRSYTWLELINPAETFGDIVIDYPYVECTSAAIQALTSFK
Gi: 541855   IISLQNAD----GGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAAIQALISFR
Gi: 1098634  LLSMRNSD----GGFATYETKRGGYLLELLNPSEVFGDIMIDYTYVECTSAVMQALRHFR
Gi: 1169548  LLNLQNIGSFEYGSFATYEKIKAPLAMETLNPAEVFGNIMVEYPYVECTDSSVLGLTYFH
                 QW-4
               *  *         *    *   **** ***  *****   *         *    *
SEQ ID NO:6  KLYPGHRREEIQHCIDKATTFIEKIQASDGSWYGSWGVCFTYGAWFGVKGLIAAGRSFSN
Gi: 4589852  KLYPGHRREEIQCCIEKAASFIEKTQASDGSWYGSWGVCFTYGTWFGVKGLIAAGKSFNN
Gi: 541855   KLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKGLVAVGKTLKN
Gi: 1098634  EYFPDHRATEIRETLNQGLDFCRKKQRADGSWEGSWGVCFTYGTWFGLEAFACMGHIYQN
Gi: 1169548  KYF-DYRKEEIRTRIRIAIEFIKKSQLPDGSWYGSWGICFTYAGMFALEALHTVGETYEN
                               QW-3
                  * ** * *   *****   *                  * *  * ** * *
SEQ ID NO:6  ---CSSIRKACEFLLSKQLPSGGWGESYLSCQNKVYSNLEGNRSHVVNTGWAMLALIDAG
Gi: 4589852  ---CSSIRKACEFLLSKQLPSGGWGESYLSCQNKVYSNVESNRSHVVNTGWAMLALIDAE
Gi: 541855   ---SPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAWAMLALIGAG
Gi: 1098634  RTACAEVAQACHFLLSRQMADGGWGEDFESCEQRRY--VQSAGSQVHSTCWALLGLMAVR
Gi: 1169548  ---SSTVRKGCDFLVSKQMKDGGWGESMKSSELHSY--VDSEKSLVVQTAWALIALLFAE
                 QW-2
                  *    *   *   *           ****  * * *  **  *** ***
SEQ ID NO:6  QAKRDSQPLHRAAAYLINSQLEDGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYQS
Gi: 4589852  QAKRDPTPLHRAAVYLINSQMENGDFPQQEIMGVFNKNCMITYAAYRNVFPIWALGEYRH
Gi: 541855   QAEVDRKPLHRAARYLINAQMENGDFPQQEIMGVFNRNCMITYAAYRNIFPIWALGEYRC
Gi: 1098634  HPDISAQ--ERGIRCLLGKQLPNGDWPQENISGVFNKSCAISYTNYRNIFPIWALGRFSS
Gi: 1169548  YPNKEV--IDRGIDLLKNRQEESGEWKFESVEGVFNHSCAIEYPSYRFLFPIKALGMYSR
                     QW-1

SEQ ID NO:6  ---QVL--QSR
Gi: 4589852  ---RVL--QSQ
Gi: 541855   ---QVLLQQGE
Gi: 1098634  LYPDNTLAGHI
Gi: 1169548  AYETHTL----
```